

Climate Change Effects on Sri Lankan Paddy Yield: An Initial Investigation Using Data Mining Algorithms

S Shanmuganathan

^aGeoinformatics Research Centre,
School of Computing and Mathematical Science,
Auckland University of Technology, New Zealand
Email: subana.shanmuganathan@aut.ac.nz

Abstract: In recent times, climate change (CC) and its effects on key crops, such as rice, wheat and maize, have drawn significant research interest alongside population increase, economic growth and changing diet patterns, all of them considered as the driving forces influencing earth's food and water ecosystems. Despite recent technological advances, such as from improved plant breeds (*cultiva* or varieties) to irrigation systems, which have contributed towards improving the world's staple food production significantly, climate still remains as the key factor in agricultural productivity. Hence, understanding the effects of climate change on various staple food crops has become the utmost priority in many aspects especially, to overcome the threats to the world food security. As a result of this, many institutions concerned over related issues and research communities, in recent years, have turned their focus into modelling the phenomenon at various scales and levels. Contemporary research on modelling the climate change patterns in weather conditions and their effects is summarised. Existing crop forecasting models vary significantly in spatiotemporal scales and levels, the lowest being at the micro (e.g. the field or farm at specific days/weeks), and the highest at macro (e.g. regional /district, at months/years) or global, the crops being studied include, staple crops (maize, rice and wheat) and vineyards all using an array of variables characterised by 1) historic (using a multitude of sources, e.g. metrological, phenological, satellite imagery and wireless sensors at the micro scale, or 2) simulated data (www.ncdc.noaa.gov/oa/ncdc.html), both against observed yield

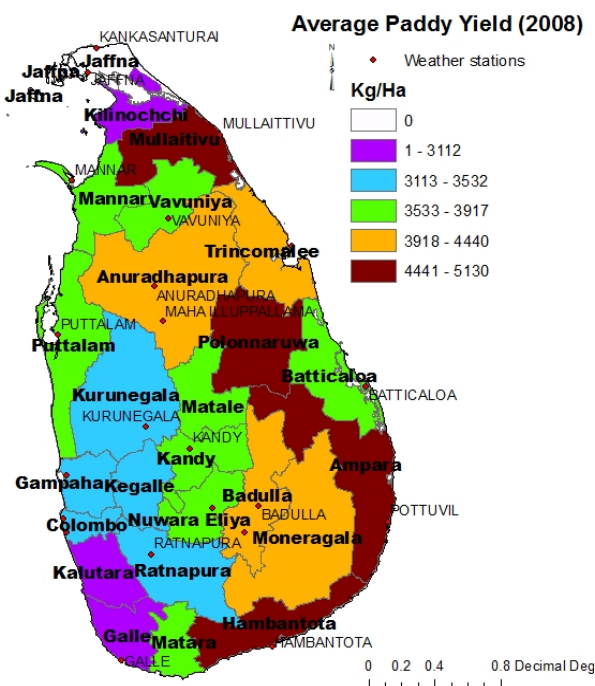


Figure 1. Paddy regions (in **bold** type) and weather stations (red spots). Source: www.statistics.gov.lk and www.ncdc.noaa.gov/cdoweb

In this context, the paper presents an initial investigation in which four data mining algorithms are explored to analyse the rice crop data in Sri Lankan administrative divisions, as an example study. Rice is the main staple food for Sri Lankans and paddy cultivation in the country dates back to as early as 800 BC. Presently, paddy is being cultivated as a wetland crop, either rain fed or irrigated. Lately, the country's estimated total land under cultivation is said to be approximately 708,000 Hectares cultivated in two seasons "Maha" and "Yala", that correspond to the country's two monsoons, North-east monsoon (from September to following March) South-west (from May to end of August). Paddy yield in various Sri Lankan divisions is presented in figure 1 based on 2008 average annual production obtained from www.statistics.gov.lk

The results of this investigation reveal interesting correlations between recent climate and paddy yield in nine regional divisions of Sri Lanka over "Yala" paddy season despite the gaps in the climate data that cannot be analysed using geostatistical or conventional methods due to the gaps in the data.

Keywords: Soil nutrient, grapevine, geostatistical analysis

1. INTRODUCTION

In recent times, climate change (CC) and its effects on key crops, such as, rice, wheat and maize, have drawn significant research interest alongside population increase, economic growth and changing diet patterns, also described as the driving forces influencing earth's food and water ecosystems (Liu, et al., 2013). Despite the recent advances, e.g. from improved crop varieties to better management practices that have contributed significantly towards improving various staple crop yields (Papademetriou, 2000), climate still remains the key factor in agricultural productivity (Rosenzweig, et al., 1993). On the other hand, potential climate change effects are predicted to have far-reaching implications on humanity if proper measures are not adapted to counterbalance the change effects more specifically, in regions that are most vulnerable to drastic climate change effects and food deficits (Fraser, 2013). In view of the current situation, to overcome the threats to world food security, many institutions both academic and research, have embarked on modelling the CC effects on the key crops at various spatiotemporal scales and some such major efforts are outlined in section 2. Section 3 gives a background on the divisions paddy is cultivated in Sri Lanka and data used in this study. Section 4 elaborates on the methodology. In the final section, preliminary results of this research on the recent CC effects on paddy yield in the 9 of the 25 administrative divisions of Sri Lanka (figure. 4) are discussed

2. MODELLING CLIMATE CHANGE EFFECTS ON CROPS

2.1. Climate-crop modelling efforts

Based on recent research by Socioeconomic Data and Applications Center (**SEDAC**), on a world scale, the agricultural sector is facing many challenges due to growing global populations, land degradation, and loss of cropland to urbanization. Despite the advances achieved in food production to keep in pace with population growth, recently the world has seen many serious regional deficits, and poverty related nutritional deficiencies. Yet again climate change is seen as the key factor that could affect food production and availability in many parts of the world, particularly those most prone to drought and famine. SEDAC, one of the Distributed Active Archive Centers (DAACs) in the Earth Observing System Data and Information System (EOSDIS) of the U.S. National Aeronautics and Space Administration (<http://sedac.ciesin.columbia.edu/maps/services#crop-climate>) focuses on human interactions in the environment. Its mission is to develop and operate applications that support the integration of socioeconomic and Earth science data and to serve as an "Information Gateway" between the Earth and social sciences.

The Agricultural Model Intercomparison and Improvement Project (**AgMIP**), a major international effort is aimed at linking the climate, crop, and economic modeling communities with cutting-edge information technology to produce improved crop and economic models and the next generation of climate impact projections for the agricultural sector. Its "mission is to improve substantially the characterization of world food security due to climate change and to enhance adaptation capacity in both developing and developed countries" <http://www.agmip.org/>. The project consisting of several teams, one each for Climate, Crop Modelling, Economics and IT looks at issues related to climate change and its effects on key crops divided into 1) Regional Assessments, 2) Global Assessments, 3) Model Improvement, 4) Model Intercomparisons and 5) Methods. The effort is supported by state as well as academic institutions.

CGIAR is a global agriculture research partnership focused on securing food for the future. It consists of 15 research centres who are members of the CGIAR which is in collaboration with hundreds of partner organizations (www.cgiar.org). The CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS) is aimed at addressing the increasing challenge of global warming and declining food security on agricultural practices, policies and measures through a strategic collaboration between CGIAR and Future Earth. This is in view of the fact that "No single research institution working alone can address the critically important issues of global climate change, agriculture and food security" (http://ccafs.cgiar.org/about#UkOm9YZ_NBk).

MACSUR describes itself as "a knowledge hub within FACCE-JPI (Joint Programming Initiative for Agriculture, Climate Change, and Food Security)" <http://www.macsur.eu/index.php/about-macsur>. The FACCE-JPI has a Scientific Research Agenda defined with five core research themes to address the impacts of climate change on European agriculture. MACSUR is aimed at gathering the excellence of existing research in livestock, crop, and trade science to describe how climate variability and change will affect regional farming systems and food production in Europe in the near and the far future, and the associated risks and opportunities for European food security.

With that short introduction on a few international efforts on the theme, in the next section, a few climate-rice models are outlined.

2.2. Climate-rice models

Using tiny endophytic fungi, a new strategy is being researched for mitigating the effects of climate change on plants in natural and agricultural ecosystems. This emerging area of research called "syntrophics" for symbiosis-altered gene expression does not change the DNA of the host in this case, the rice plant. Instead, a naturally occurring phenomenon is recreated by colonising the fungi into the rice plants to confer stress tolerance. With rice yields projected to decrease by 15 per cent in developing countries by 2050, this is seen as a pragmatic way forward as the fungi also increased seed yields and root systems in rice. The fungi species that exist naturally within native coastal (salt-tolerant) and geothermal (heat-tolerant) plants was suggested to be useful in adapting plants to drought, salt and temperature stressors predicted to worsen in future years due to climate change (Rodriguez, 2011)

In (Tazhibayeva & Townsend, 2012), a three-stage production function for rice cultivation is presented. The function incorporates the sequential nature of both production shocks, including weather and input choices based on sequentially updated information sets of history of realised shocks and observed changes in crop growth. The research was focused on a detailed partial equilibrium analysis of the effects of climate change on yields at the crop-plot level for modelling the global general equilibrium of environmental changes. The authors integrated an economic model of rice production with soil science crop growth modelling, weather simulators, and global climate change models.

It is shown in (Iizumi, et al., 2013) that moderate-to-marked yield loss over a substantial percentage (26–33%) of the harvested area of key crops as reliably predictable if climatic forecasts were near perfect. Interestingly, rice and wheat production were stated to be reliably predictable at three months before the harvest using within-season hind casts. Even though, the reliabilities of estimates were said to have varied substantially by the crop, rice and wheat yields were found to be the most predictable, followed by soybean and maize. The reasons stated for the variation in the reliability of the estimates were the differences in crop sensitivity to the climate and the technology used by the crop-producing regions. More importantly, the

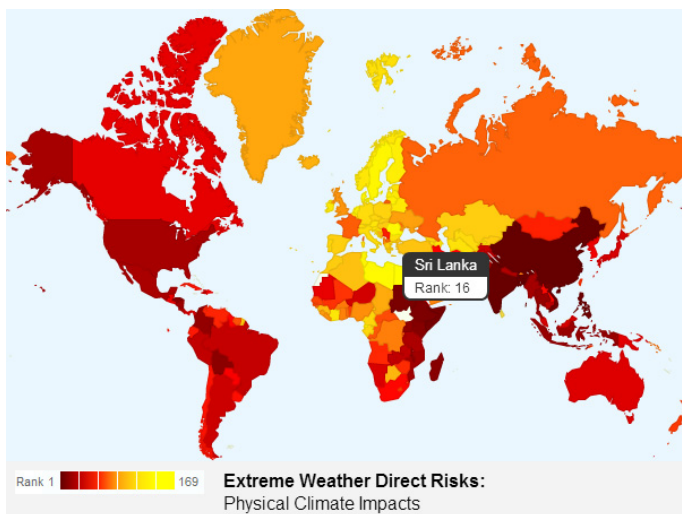


Figure 2. The world map displaying the different country rankings for extreme weather direct risks calculated based on information on human losses from natural disasters in 222 countries since 1900. Among the disaster categories tracked by EM-DAT <http://www.emdat.be/>, five were considered in the study as relevant for climate change analysis and they are: floods, droughts, extreme heat, wind storms and wild fires. Countries in the map are coloured in descending order of impact, from dark red to yellow. Using the online interactive map <http://international.cgdev.org/page/mapping-impacts-climate-change-country-rankings-for-four-different-dimensions-of-climate-impact>: Extreme Weather, Sea Level Rise, Agricultural Productivity Loss and Overall could be established. Source (Wheeler, 2011)

findings of that research revealed that the use of seasonal climatic forecasts to predict crop failures to be useful for monitoring global food production as it could be used to encourage the adaptation of food systems to climate extremes.

In a different kind of assessment presented in (Watson, et al., 2013) each ecoregion's adaptive capacity was studied based on spatial analysis of its natural integrity, the proportion of intact natural vegetation to present a measure of global ecosystem vulnerability. The research reveals that the relationship between intactness (adaptive capacity) and stability (exposure) as widely varying across the ecoregions. Based on this measure some of the ecoregions located in southern and southeastern Asia, western and central Europe, eastern South America and southern Australia are stated to be most vulnerable. The paper as well provides a matrix that highlights the potential implications of this vulnerability assessment for adaptation planning and offers a spatially explicit management guide to ensure the applicability of these findings to conservation.

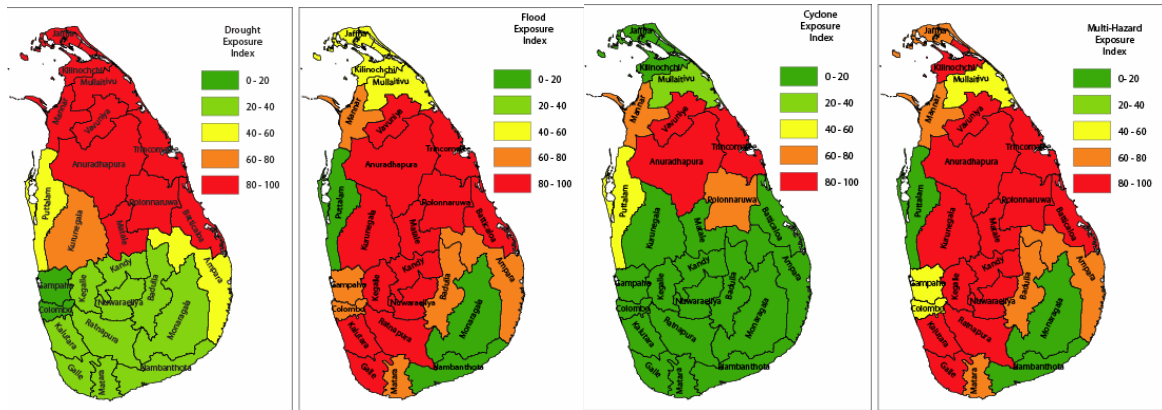


Figure 3. Maps showing a pilot level climate change Vulnerability Index for exposure arising from left to right drought, flood, cyclone and multi hazard mapped at district level. Some farming districts, such as Nuwara Eliya, Badulla, Moneragala, Ratnapura and Anuradhapura are the most vulnerable to CC due to their heavy reliance on primary agriculture. Source (Eriyagama, et al., 2010).

Meanwhile, in the Sri Lankan context (Eriyagama, et al., 2010) developed a vulnerability index for difference risk exposures (figure 3). In view of the statements and findings from the above research efforts, this paper looks at unravelling useful correlations between recent climate and rice (paddy) yield in the Sri Lankan context using data mining algorithms.

3. THE METHODOLOGY

The methodology consists of three sections, namely 1) climate data, 2) paddy yield data and 3) yield classes.

3.1. Climate data

Monthly temperature average calculated using the daily global temperature data obtained from www.ncdc.noaa.gov/cdoweb for the respective weather stations relating to the divisions (figure 1) is analysed in this research. Even though long term annual paddy yield data was available the weather data was not which impedes the use of conventional methods for this purpose.

3.2. Paddy yield data

Paddy yield from 1979-2008 for the “Yala” season obtained from Sri Lanka statistics (www.statistics.gov.lk) is used in this study. The “Yala” paddy yield and monthly maximum, minimum and average temperatures for the growing season (May-Aug) data combined into one file was used to generate classification /decision tree models rules with C5.0, CRT (classification and regression tree), CHARD and QUEST functions in SPSS.

In addition, an ANN model as well was tested for predicting “Yala” paddy yield using the same data set. Due to inconsistencies in the temperature data, this initial study has been limited to 9 divisions and over the time slots historic data was made available. The divisions included in the analysis are: Anuradhapura /ANU (1983-2008), Batticaloa/BA (1979-2008), Colombo/COL (1979-2008), Galle (2002-2008), Hambantota/HAM (1979-2008), Jaffna/JA (1979-1983), Kandy/KA (2007-2008), Mannar/MA (1982-1997) and Vavuniya/VA (2003) The other major issue experienced in this study is the missing values /gaps in the weather data.

3.3. Yield classes

The C5.0 and QUEST algorithms require the “output” variable to be “categorical” and the “input” fields/variables can be numeric ranges. Hence the “Yala” yield data is converted into three classes (“low” <3250 kg ha⁻¹, “med” or medium 3250 - 4250 kg ha⁻¹ and “high” >4250 kg ha⁻¹) for use as “target” based on the yield data distribution (figure 4). The input variables used are May-August month minimum, maximum and average temperatures.

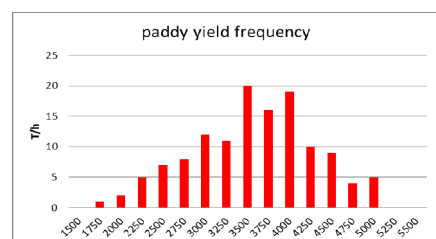


Figure 4. Data distribution of “Yala” yield (Kg/Ha) in 1979-2008

4. RESULTS AND DISCUSSION

Results obtained from the different data mining algorithms and the ANN model are discussed in the section.

4.1. Data mining algorithm results

Table 1. Confusion matrices of the different data mining algorithms along with respective rules.

C5

class	high	low	med	rules
high	11	0	7	Rule 1 (7; 1.0) if district in ["HAM"] and maxT_jul <= 34.5 then high* Rule 2 (3; 0.333) if district in ["MANNAR"] then high
low	5	25	8	Rule 1 (4; 1.0) if district in ["ANU"] and minT_jul <= 23.39 then low
med	10	5	47	Rule 1 (9; 0.667) if district in ["ANU"] and minT_jul > 23.39 then med Rule 3 (4; 0.75) if district in ["HAM"] and maxT_jul > 34.5 then med

Chi-square= 65.29 dt=4 probability=0

CRT

class	high	low	med	rules
high	12	0	6	Rule 1 (7; 1.0) if district in A and maxT_jul <= 34.64 then high * Rule 2 (1; 1.0) if district in A and maxT_jul > 34.64 and minT_may <= 24.25 then high Rule 3 (1; 1.0) if district in A and aveT_jul > 23.25 and maxT_jul <= 33.945 and minT_aug > 26.445 then high Rule 2 (5; 1.0) if district in A and aveT_jul <= 23.25 and maxT_jun > 31.5 then low
low	1	29	8	Rule 3 (2; 1.0) if district in A and aveT_jul > 23.25 and maxT_jul > 33.945 and maxT_jun <= 33.25 then low
med	12	11	39	Rule 1 (3; 1.0) if district in A and maxT_jul > 34.64 and minT_may > 24.25 then med Rule 2 (1; 1.0) if district in A and aveT_jul <= 23.25 and maxT_jun <= 31.5 then med Rule 3 (12; 1.0) if district in A and aveT_jul > 23.25 and maxT_jul <= 33.945 and minT_aug <= 26.445 then med ** Rule 4 (15; 0.6) if district in A and aveT_jul > 23.25 and maxT_jul > 33.945 and maxT_jun > 33.25 then med ****

Chi-square=64.714 df=4 probability=0 A= ["ANU" "BATT" "HAM" "JAFF" "KA" "MANN"]

CHAID

class	high	low	med	rules
high	11	0	7	Rule 1 (7; 1.0) if district in ["HAM"] and maxT_jul <= 34.72 then high Rule 2 (1; 1.0) if district in ["HAM"] and maxT_jul > 34.72 and minT_aug <= 23.61 then high
low	0	21	17	
med	13	1	48	Rule 2 (3; 1.0) if district in ["HAM"] and maxT_jul > 34.72 and minT_aug > 23.61 then med

Chi-square=68.615 df=4 probability=0

QUEST

class	high	low	med	rules
high	14	0	4	
low	5	24	9	Rule 1 (2; 1.0) if district in ["ANU" "BATT" "COL" "KA"] and minT_jul <= 22.90 then low
med	21	3	38	Rule 1 (17; 0.882) if district in ["ANU" "BATT" "COL" "KA"] and minT_jul > 22.90 then med ***

Chi-square=64.802 df=4 probability=0

Of the data mining results (Table 1), CHAID algorithm produced the highest chi-square 68.615. For the classes, in the confusion matrix “high”, “low” and “med” studied QUEST, CRT and CHAID algorithms gave the highest correct predictions respectively.

Based on the different algorithm rules the following could be established between “Yala” paddy yield and May to August maximum, minimum and average temperatures;

- 1) July maximum temperature in "ANU" "BATT" "HAM" "JAFF" "KA" "MANN" districts/ divisions has to be $\leq 34.5^{\circ}\text{C}$ for “Yala” paddy yield to be high. This is true 7 instances with 1.00 confidence interval (Table 1 *).
- 2) Again in “ANU” “BATT” “HAM” “JAFF” “KA” “MANN” divisions three conditions, July average > 23.25 , July maximum temperature ≤ 33.945 and August minimum temperature ≤ 26.445 has led to medium “Yala” paddy yield met in 12 instance with 1.0 confidence interval (Table 1 **)
- 3) In divisions "ANU" "BATT" "COL" "KA" July minimum temperature > 22.90 leads to medium “Yala” yield. This is met in 17 instance at 0.882 confidence interval (Table 1 ***)
- 4) In ANU" "BATT" "HAM" "JAFF" "KA" "MANN" July average temperature > 23.25 , maximum temperature > 33.945 and June maximum temperature > 33.25 have led to medium “Yala” paddy yield met in 15 instances at 0.6 confidence interval (Table 1 ****)

As it appears, Anudhapura, Batticaloa, Hambantota, Jaffna, Kandy and Mannar divisions have been affected by subtle increases in July maximum temperature. Meanwhile, Anudhapura, Batticaloa, Colombo and Kandy as well have been affected by increases in July minimum temperature even though some Sri Lankan research show that the southern parts of the island would not be affected by droughts.

4.2. ANN results

Based on ANN prediction model results (Table 2), of all the monthly temperatures, July maximum temperature contributes the most i.e., 0.51 to the “Yala” yield in the predictions. This is followed by maximum temperatures in August and May and then by the August average temperature, the contributions by the latter variables are between 0.06-0.04. The accuracy of the predictions was 88.09 per cent. Evaluation graphs (figure 5) show the ANN training and testing accuracies.

5. CONCLUSIONS

The paper looked at the major research efforts undertaken by international, academic and national institutions, and communities to understand the climate change effects on key crops to adapt proper measures and cultivation practices to suit the changing climate for securing food for the ever increasing world population. The results of this study show that even subtle increases in July maximum and minimum temperatures have a negative impact on the “Yala” paddy yields in most of the divisions namely, Anudhapura, Batticaloa, Hambantota, Jaffna, Kandy and Mannar.

Table 2. ANN prediction results

Analysis

Estimated accuracy: 88.09
 Input Layer: 22 neurons
 Hidden Layer 1: 3 neurons
 Output Layer: 1 neurons
 Relative Importance of Inputs

district	0.511638
maxT_jul	0.067507
maxT_aug	0.067349
maxT_may	0.0654055
aveT_aug	0.0448963
minT_jul	0.0411945
aveT_may	0.0364861
minT_jun	0.0267123
aveT_jun	0.02117
aveT_jul	0.0183511
minT_may	0.0158729
maxT_jun	0.0145307
minT_aug	0.00631416

Target: Yala

Build Settings

Use partitioned data: false
 Method: Quick
 Stop on: Default
 Set random seed: true
 Set random seed: true
 Prevent overtraining: true
 Sample %: 50.0
 Optimize: Memory

Training Summary

Model type: Neural net
 Stream: Stream1
 User: ss Date built: 25/09/13 00:14
 Application: Clementine 10.1
 Elapsed time for model build:
 0 hours, 0 mins, 1 secs

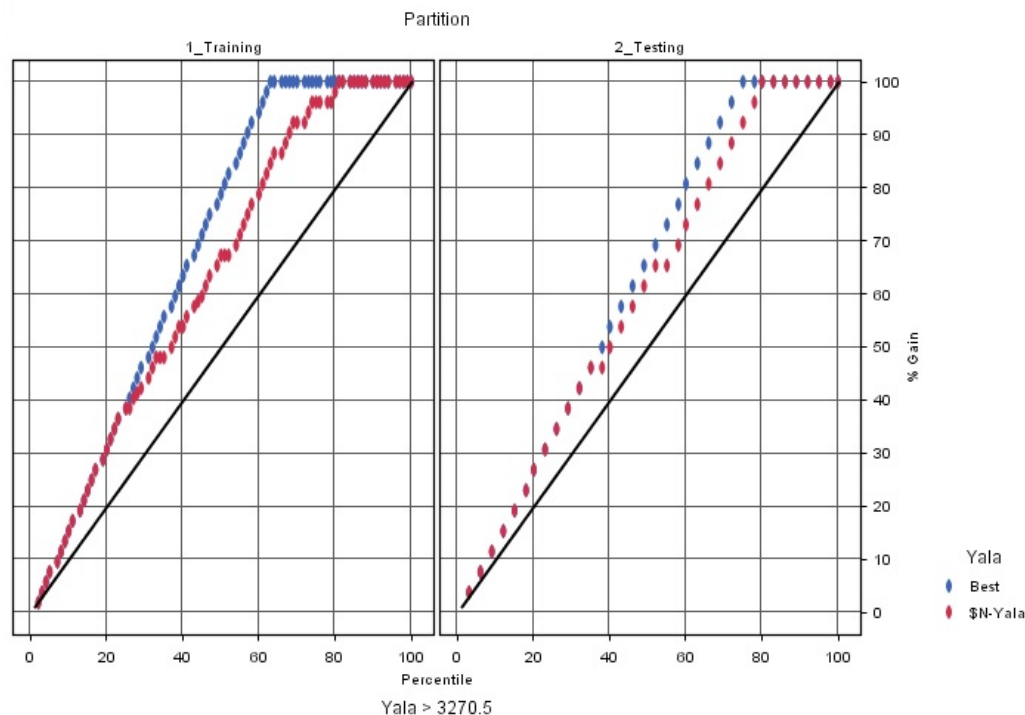


Figure 5: ANN evaluation graphs showing the “Yala” paddy yield predictions

REFERENCES

- Papademetriou, M. K., 2000. in [Editors] Papademetriou, Minas K.; Dent, Frank J; Edward , M. H “Bridging the Rice Yield Gap in Asia and the Pacific. Bangkok, Thailand, 5-7 October, 1999. , presented at the Expert Consultation on “Bridging the Rice Yield Gap in Asia and the Pacific”, Bangkok, Thailand, 5-7 October, 1999. The Consultation was organized and sponsored by the FAU, UN.
- Eriyagama, N., Smakhtin, V., Chandrapala, L. & Fernando, K., 2010. *Impacts of climate change on water resources and agriculture in Sri Lanka: a review and preliminary vulnerability mapping*. Colombo, Sri Lanka, (IWMI Research Report 135). [doi:10.5337/2010.211]: International Water Management Institute, 51p. <http://www.iwmi.cgiar.org/publications/iwmi-research-reports/iwmi-research-report-135/>.
- Fraser, E. D. G., Simelton, E., Termansen, M., Gosling, S. N. and South, A., 2013. “Vulnerability hotspots”: Integrating socio-economic and hydrological models to identify where cereal production may decline in the future due to climate change induced drought. *Agricultural and Forest Meteorology* 170: 195-205, Issue <http://dx.doi.org/10.1016/j.agrformet.2012.04.008>.
- Iizumi, T. et al., 2013. Prediction of seasonal climate-induced variations in global food production. *Nature Climate Change* 3, 904–908 (2013) doi:10.1038/nclimate1945.
- Liu, J. et al., 2013. A Global and Spatially Explicit Assessment of Climate Change Impacts on Crop Production and Consumptive Water Use. *PLOS ONE*, February 2013 | Volume 8 | Issue 2 | e57750 (www.plosone.org), p. 13.
- Rodriguez, R., 2011. *Climate adaptation of rice*. [Online] Available at: <http://phys.org/news/2011-07-climate-rice.html#jCp> [Accessed 27 Sep 2013].
- Rosenzweig, C., Parry, M. L., Fischer, G. & Froberg, K., 1993. *Climate change and world food supply. Research Report No. 3.* , Oxford: University of Oxford, Environmental Change Unit. 1993. www.ciesin.org/docs/004-046/004-046.html.
- Tazhibayeva, K. & Townsend, R. M., 2012. *The Impact of Climate Change on Rice Production: Heterogeneity and Uncertainty*, http://robertmtownsend.net/sites/default/files/files/papers/working_papers/ImpactofClimateChangeonRiceYieldsDec2012.pdf.
- Watson, J. E. M., Iwamura, T. & Nathalie, B., 2013. Mapping vulnerability and conservation adaptation strategies under climate change. *Nature Climate Change*, Volume <http://dx.doi.org/10.1038/nclimate2007>, pp. 1758-6798.
- Wheeler, D., 2011. *Quantifying Vulnerability to Climate Change: Implications for Adaptation Assistance - Working Paper 240*, s.l.: <http://international.cgdev.org/publication/quantifying-vulnerability-climate-change-implications-adaptation-assistance-working>.